

Busy beavers gone wild

Grégory Lafitte

Laboratoire d'informatique fondamentale de Marseille, CNRS,
Aix-Marseille Université, 39, rue F. Joliot-Curie,
13453 Marseille Cedex 13, France *
Gregory.Lafitte@lif.univ-mrs.fr

We show some incompleteness results *à la* Chaitin using the busy beaver functions. Then, with the help of ordinal logics, we show how to obtain a theory in which the values of the busy beaver functions can be provably established and use this to reveal a structure on the provability of the values of these functions.

1 Introduction

When considering a class of Turing machines that halt when started from a blank tape, *busy beavers* for that class, as coined by Tibor Radó, are those Turing machines which eventually halt after the maximum number of steps (or after producing the maximum number of non-blank symbols). Finding for a class of Turing machines the busy beaver champion (for any of the two competitions) is of course uncomputable [19]. Nevertheless, efforts have been made to try to find the champions starting with the smallest classes of Turing machines. Finding all the different kinds of behavior of Turing machines in a class is quite exhilarating an endeavor. Rona J. Kopp [9] and Allen H. Brady [3] have thoroughly studied the four states and two symbols class and Heiner Marxen and others [15, 16] have tried for more than two decades to experimentally find the busy beaver champion for the five states and two symbols class. Heiner Marxen found a candidate but no one has been able to prove that no other machine *does better*. This is most certainly because the behaviors involved are too complicated. This state of affairs is an encouragement to the study of what we have called *low-level reverse computability*: What does it take to be able to comprehend the behavior of a certain class of Turing machines? What is needed to be able to prove these behaviors? These questions raise the links between the study of busy beavers and incompleteness.

To start off, there is a need to see the strong links between computability and incompleteness. For instance, one can obtain a first form of the first incompleteness theorem by considering propositions of the form $n \notin X$, where X is a non-recursive but recursively enumerable set, *e.g.*, the *diagonal halting* set \mathcal{H} . Even if the language of the considered theory does not contain \in , there is a simple algorithm that generates given n the proposition “ $n \notin X$ ”. Given a sound (every provable statement is true) recursively enumerable theory T , there is a number n_0 such that $n_0 \notin X$ but T does not prove it. The proof is direct: Suppose that there is no such n_0 , then we would have that T proves “ $n \notin X$ ” if and only if $n \notin X$, and X would be recursive (generate the theorems of T and at the same time enumerate X ; if $n \in X$ then n will eventually show up in the enumeration; otherwise, “ $n \notin X$ ” will eventually show up in the theorems of T and be true by the soundness assumption). We thus have a true sentence, “ $n_0 \notin X$ ”, which is not provable in T .

*We acknowledge the support of the French National Research Agency Sycomore grant.

Incompleteness is also famously linked to computability via Chaitin's incompleteness theorem. Chaitin's result, showing that there are unprovable statements on Kolmogorov-Chaitin complexity¹, is a form of Gödel's first incompleteness theorem. Actually, Kolmogorov showed in the sixties that the set of random (or incompressible) numbers, *i.e.*, $\{x : K(x) \geq x\}$, is recursively enumerable but not recursive, and, by the above argument, this is already a version of Gödel's first incompleteness theorem. Moreover, Kolmogorov's proof can be seen as an application of Berry's paradox ("the least integer not nameable in fewer than seventy characters" has just now been named in sixty-three characters). Following Boolos [1], it is thus no wonder that we can get proofs using this Kolmogorov complexity function (or other *similar* computability-related functions) of both incompleteness theorems (see [10]).

One of the reasons of the existence of the quest for better understanding the incompleteness phenomenon lies in the peculiarity of Gödel's unprovable statements. They are not natural mathematical statements: no mathematician has ever stumbled on them (or should we say *over them*?). And thus, it seems to many that normal mathematical practice is not concerned with the incompleteness phenomenon. More and more results show however the contrary. In particular, Harvey Friedman's Π_1^0 statements, that are unprovable in Zermelo-Fraenkel (ZF) set theory and need the 1-consistency of strong set-theoretical unprovable statements (going way beyond ZF) to be proved, are examples of such results.

Nevertheless, incompleteness theorems only provide unprovable statements like the *consistency* of a theory, that are of an unclear nature. What combinatorial properties does the consistency statement bring to a theory? Having a link between consistency (or soundness) and computability, in particular Kolmogorov complexity, would make possible an understanding of what properties consistency adds to a theory. Adding consistency as an axiom would then yield new combinatorial properties because of the existing links between combinatorics and Kolmogorov complexity. For a detailed exposition of the links between Gödel incompleteness and Kolmogorov complexity, see [10].

In the first part of this article, we start by giving an incompleteness theorem *à la* Chaitin using the busy beaver functions and then make precise the kind of incompleteness theorems one can obtain with them. In the second part, we use ordinal logics to construct a theory in which the values of the busy beaver functions can be provably established and use this to find a structure on the provability of the values of these functions.

Notational conventions

On top of the usual logical connectives (\wedge , \vee and \neg), we will respectively denote the logical connectives of *implication* and *equivalence* by \supset and \equiv .

When we have a recursively enumerable axiomatic theory T , Cons_T designates the arithmetical sentence that there is no proof of $0 = 1$. When the axioms of T are defined by a formula ϕ , $\text{Cons}(\phi)$ will

¹Loosely speaking, the Kolmogorov-Chaitin complexity of a natural number n , denoted by $K(n)$, is the smallest size of a program which generates n .

Another useful variant of Kolmogorov-Chaitin complexity can be defined as follows. To each enumeration $\{\varphi_i\}_{i \in \mathbb{N}}$ of computable partial functions, we associate a variant of K : $K_\varphi(x|y) = \text{smallest } e \text{ such that } \varphi_e(y) = x$, and $K_\varphi(x) = K_\varphi(x|0)$.

Chaitin's version of Gödel incompleteness is the following theorem.

Theorem 1.1 (Chaitin's incompleteness). *Let T be an arithmetical-able consistent theory. There is a constant c_T such that for all x , " $K(x) > c_T$ " is unprovable in T .*

For more on Kolmogorov-Chaitin complexity, the reader is referred to [14]. For a formal definition of *arithmetical-ability*, see [10].

designate Cons_T using that formalization. The provability predicate is designated by Prov_T .

The partial recursive functions computed by Turing machines, following a fixed convention, are denoted by $\{\varphi_i\}_{i \in \mathbb{N}}$ (agreeing with the Turing machines' coding: φ_i is the partial function *computed* by T_i). The sets $\{W_i\}_{i \in \mathbb{N}}$ denote the recursively enumerable sets, *i.e.*, the domains of partial recursive functions.

Concerning computability, the reader is referred to [17, 18, 21, 20, 23, 25].

2 Incompleteness and busy-beaverhood

In the sixties, Tibor Radó, a professor at the Ohio State University, thought of a simple non-computable function other than the standard halting problem for Turing machines. Given a fixed finite number of symbols and states, select those Turing machines which eventually halt when run with a blank tape. Among these programs, find the maximum number of non-blank symbols left on the tape when they halt. Alternatively, find the maximum number of time steps before halting. These functions are well-defined but uncomputable. Tibor Radó called them the busy beaver functions. For more on the busy beaver problem, read [19, 11, 13, 2, 9, 3, 5, 6, 4, 16, 15, 12].

Alternative functions can be defined that are close in nature to these busy beaver functions. Let σ^{steps} be the function which to i gives the maximum number of steps for which a Turing machine with code $\leq i$ will keep running before halting, starting with a blank tape. For a Turing machine M , t_M denotes the time complexity function of M : $t_M(x) = s$ if $M(x)$ halts after s steps. Following the busy beaver functions' definitions, we define σ^{value} to be the function which to i gives the maximum number which a Turing machine with code $\leq i$ will output, following a fixed convention, after halting starting with an input $\leq i$. These functions are in a sense inverses of the K_φ function (see footnote 1).

Other functions can be defined following classical Kolmogorov complexity, *e.g.*, the function which to n gives the biggest number with Kolmogorov complexity lower than n .

We call these functions the σ functions. For each variant, we can define a function focusing on maximizing the *number of steps*, *e.g.*, σ^{steps} , or the *output values*, *e.g.*, σ^{value} . The value of either of these functions on a certain x is computable from x and the value of the other function on input $x + c$ for a certain constant c (see [22]).

A result similar to Chaitin's incompleteness result concerning Kolmogorov complexity can be obtained concerning the σ functions²:

Theorem 2.1 (Chaitin-like incompleteness theorem for σ functions). *Let σ be one of the σ functions. Let T be an arithmetical-able³ consistent theory. There is a constant n_T^σ such that*

$$T \vdash \text{Cons}_T \supset \forall s \neg \text{Prov}_T(\ulcorner \sigma(n_T^\sigma) < s \urcorner). \quad (1)$$

Proof. Consider a Π_1 formula ϕ_σ in the language of T such that $\phi_\sigma(x, s)$ expresses that $\sigma(x) < s$.

Working in T , for a given x , take the smallest s such that $\text{Prov}_T(\ulcorner \phi_\sigma(x, s) \urcorner)$ holds. T being consistent and $\phi_\sigma \Pi_1$, $\phi_\sigma(x, s)$ also holds.

²A σ function is any of the busy beaver functions defined above.

³For a formal definition of *arithmetical-ability*, see [10].

$\text{Prov}_T(\ulcorner \phi_\sigma(x, s) \urcorner)$ is a Σ_1 formula and thus can be seen as $\exists y \psi(x, s, y)$ or equivalently $\exists \langle s, y \rangle \psi(x, s, y)$ where ψ is Δ_0 .

Thus there is a Turing machine computing ψ . Consider its code i_ψ (or its number of states or transitions, depending on the choice of σ). For large enough x , i.e., $x > i_\psi + c$, knowing that $\phi_\sigma(x, s)$ holds (using the computation through shifting, i.e., the constant c , between both types of σ functions), we know that $\sigma(x) < s$ and thus there is an $s' = \langle s'_1, s'_2 \rangle < s$ such that $\psi(x, s'_1, s'_2)$ holds. But for each $s' = \langle s'_1, s'_2 \rangle$ smaller than s , the statement $\neg \psi(x, s'_1, s'_2)$ is true by the minimality of s , and provable (being Δ_0). Thus we have $\neg \text{Cons}_T$. \square

We say that a statement ϕ is a *revelation for T* if ϕ is unprovable in T and its consistency relative to T (if T is consistent, so is $T + \phi$) is provable from itself in T :

$$T \vdash \phi \supset \text{Cons}_T(\phi)$$

We have the following result showing how being able to prove an upper bound for a σ function can be a revelation.

Theorem 2.2 (Serendipitous incompleteness theorem for σ functions). *Let σ be one of the σ functions. If T is consistent, then there exists a natural number τ_T^σ such that for all x , $\sigma(\tau_T^\sigma) < x$ is a revelation for T .*

Proof. Consider the Π_1 statement $\forall x \psi(x)_x$ equivalent to $\text{Cons}_{T+\phi}$.

$\psi \in \Delta_0$ and thus there is a machine M_ψ with code i_ψ such that M_ψ decides $\{x : \psi(x)_x\}$: M_ψ on input x eventually enters an acceptance state if $\psi(x)_x$, or a rejection state otherwise.

Consider another Turing machine M'_ψ which runs M_ψ successively on each natural number starting from 0 and stops and writes the counter example of ψ if the simulation of M_ψ enters a rejection state.

Let i'_ψ be the code of Turing machine M'_ψ . $\sigma(i'_\psi)$ makes the verification of $\forall x \psi(x)_x$ a Δ_0 property.

By using Kleene's recursion theorem on this previous construction, we find τ_T^σ such that knowing (or bounding) the value of $\sigma(\tau_T^\sigma)$ makes the verification of $\text{Cons}_{T+\sigma(\tau_T^\sigma) \leq x}$ a Δ_0 property. Knowing that T is consistent and assuming $\sigma(\tau_T^\sigma) \leq x$, T thus proves $\text{Cons}_T(\sigma(\tau_T^\sigma) \leq x)$.

By Gödel's second incompleteness theorem, $\sigma(\tau_T^\sigma) \leq x$ is an unprovable statement in T . \square

This can be done for other uncomputable functions, like Kolmogorov complexity functions (using the Turing-completeness of their graphs). Truth-table completeness is a plus to be able to have a finite revelation (finite number of constants $c_{i,T}^K$ giving the revelation $\bigwedge K(x) > c_{i,T}^K$).

The same idea can be used to find a constant $\tau_{T'}^\sigma$ such that if T and T' are consistent, then for all x , $T \vdash \sigma(\tau_{T'}^\sigma) < x \supset 1\text{-Cons}_{T'}$. This can be done for other soundness properties than 1-consistency.

By combining the proofs of the two previous theorems 2.1 and 2.2, we obtain :

Corollary 2.3 (Busy beaver pairs). *Let σ be one of the σ functions. If T is consistent, then there exist $s_0^{\sigma,T}$ and $s_1^{\sigma,T}$ such that for all x ,*

$$T \vdash \sigma(s_0^{\sigma,T}) < x \supset \forall s \neg \text{Prov}_T(\ulcorner \sigma(s_1^{\sigma,T}) < s \urcorner).$$

3 Having enough power to comprehend busy beavers

We start by recalling the basic notions encompassing ordinal logics. We first need the notion of *ordinal notation*.

Let $\langle \mathbf{O} \subset \mathbb{N}, \prec_{\mathbf{O}}, |\cdot| \rangle$ be defined inductively, simultaneously on $\prec_{\mathbf{O}}$ and $|\cdot| : \mathbf{O} \rightarrow \text{Ord}$, as follows:

$0 \in \mathbf{O}$ and $|0| = 0$;

If $a \in \mathbf{O}$, then $\zeta(a) = \langle a, 0 \rangle \in \mathbf{O}$, $a \prec_{\mathbf{O}} \zeta(a)$, and $|\zeta(a)| = |a| + 1$;

If φ_e is a total function which is increasing according to $\prec_{\mathbf{O}}$ (for all n , $\varphi_e(n) \prec_{\mathbf{O}} \varphi_e(n+1)$), then $\vartheta(e) = \langle e, 1 \rangle \in \mathbf{O}$, $\vartheta(e) \succ_{\mathbf{O}} \varphi_e(n)$ for all n , and $|\vartheta(e)| = \lim_{n \in \omega} |\varphi_e(n)|$.

Elements of \mathbf{O} are called *notations*. $\zeta(a)$ denotes the *successor* of a , $\vartheta(e)$ the *limit* of e . \mathbf{O} is called the *general ordinal representation system* because every *ordinal representation system* is isomorphic to $\mathbf{O}_a = \{b \prec_{\mathbf{O}} a\}$ for a particular a . Because of the definition of $\vartheta(e)$, assuming φ_e to be total and increasing, $\prec_{\mathbf{O}}$ is uncomputable, even not recursively enumerable.

Starting from a theory T , we can now define *progressions* of theories indexed by notations.

$$\begin{aligned} T_0 &= T \\ T_{\zeta(a)} &= T + \text{Cons}(T) \\ T_{\vartheta(e)} &= \bigcup_n T_{\varphi_e(n)} \end{aligned}$$

The idea is to have a sequence of theories indexed by ordinals showing *how many times* we assume the consistency of the previous theories, the predecessor theory if the ordinal is a successor and all previous theories if a limit ordinal. The difficulty lies in always having recursively enumerable axiomatic theories, *i.e.*, having a sequence of formulae ϕ_a defining the axioms of T_a and definable from previous ϕ 's. Turing [24] and later Feferman [7] came up with this idea of recursive progressions of theories. Kleene's second recursion theorem shows the existence of such recursive progressions. Notice that assuming the soundness of the base theory T , if we add its consistency to the axioms of T , then we know that this enhanced theory is also sound. Hence, soundness is inherited by consistency extensions.

Now, the problem is that T_a is not consistent for all a even when assuming the soundness of T : by Kleene's recursion theorem, there is e such that $\varphi_e(0) = \zeta(\vartheta(e))$ and for such e , $T_{\vartheta(e)}$ proves all that $T_{\zeta(\vartheta(e))}$ proves, in particular $\text{Cons}_{T_{\vartheta(e)}}$; thus $T_{\vartheta(e)}$ is inconsistent.

To ensure that T_a is a consistency extension of T , one has to make sure that a belongs to \mathbf{O} . One way of doing that is working only with a 's from a *branch* of \mathbf{O} , that is a linearly ordered subset of \mathbf{O} closed under $\prec_{\mathbf{O}}$. A better way is considering only notations, elements of \mathfrak{O} , such that if $a \in \mathfrak{O}$ and an arithmetically definable variant of " $b \in \mathbf{O}$ " is provable in T_a then $b \in \mathfrak{O}$.

In theorem 3.1 we use arguments along the lines of Turing's [24] as they were described by Feferman [7, 8].

Theorem 3.1 (Point-wise provability of σ values). *Let σ be one of the σ functions. For any consistency progression, for all $a \in \mathbf{O}$ and for all $x \in \mathbb{N}$, there exists $a_x \in \mathbf{O}$ with $a \prec_{\mathbf{O}} a_x$ such that $|a_x| = |a| + \omega + 1$ and T_{a_x} proves all true statements of the form " $\sigma(x) < \dots$ ".*

Proof. Consider a Π_1 formula $\phi_{\sigma}(x, s) = \forall t \psi_{\sigma}(t, x, s)$ in the language of T such that $\phi_{\sigma}(x, s)$ expresses $\sigma(x) < s$ and ψ_{σ} is Δ_0 .

Let σ_x be the true value of $\sigma(x) + 1$. By Kleene's recursion theorem, let e_x be such that provably in PA, for every n ,

$$\varphi_{e_x}(n) = \begin{cases} \zeta^n(a) & \text{if } \ulcorner \psi(\ulcorner i \urcorner, \ulcorner x \urcorner, \ulcorner \sigma_x \urcorner) \urcorner \text{ is true for every } i \leq n, \\ \zeta(\vartheta(e_x)) & \text{otherwise.} \end{cases}$$

Since $\forall t \psi(t, x, s)$ is true, $\vartheta(e_x) \in \mathbf{O}$ and $|\vartheta(e_x)| = |a| + \omega$. Let $a_x = \zeta(\vartheta(e_x))$. If $\neg \forall t \psi(t, x, s)$, then $T_{\varphi_{e_x}(n)}$ is T_a for all sufficiently large n and so $T_{\vartheta(e_x)}$ also proves the consistency of $T_{\vartheta(e_x)}$. Thus by Gödel's second incompleteness theorem, $T_{\vartheta(e_x)}$ is inconsistent. Hence, since we can prove in T_{a_x} the consistency of $T_{\vartheta(e_x)}$, $\forall t \psi(t, x, s)$ is also provable in T_{a_x} . \square

For any class of machines, it is thus possible with a combination of theories from a progression of theories to comprehend the busy beaver functions on that class.

By an extensive use of Kleene's recursion theorem and by varying the theory in the use of corollary 2.3, we get the following corollary.

Corollary 3.2 (Busy beaver relationships). *Let σ be one of the σ functions. If T is consistent, then for all $n > 1$, there exist $\{\mathfrak{s}_i^{\sigma, T}\}_{i < n}$ such that for all s , $\sigma(\mathfrak{s}_0^{\sigma, T}) < s$ is not provable in T , and for all x and $i < (n - 1)$,*

$$T \vdash \sigma(\mathfrak{s}_i^{\sigma, T}) < x \supset \forall s \neg \text{Prov}_T(\ulcorner \sigma(\mathfrak{s}_{i+1}^{\sigma, T}) < s \urcorner).$$

References

- [1] George Boolos. A new proof of the Gödel incompleteness theorem. *Notices of the American Mathematical Society*, 36:383–391, 1989.
- [2] Allen H. Brady. The conjectured highest scoring machines for Rado's $\sigma(k)$ for the value $k = 4$. *IEEE Transactions on Elec. Comput.*, EC-15:802–803, 1966.
- [3] Allen H. Brady. The determination of the value of Rado's noncomputable function $\sigma(k)$ for four-state Turing machines. *Mathematics of Computation*, 40:647–665, 1983.
- [4] Allen H. Brady. *The Universal Turing Machine: A Half-Century Survey*, chapter The Busy Beaver Game and the Meaning of Life, pages 259–277. Oxford University Press, Oxford, England, 1988.
- [5] A. K. Dewdney. Computer recreations: A computer trap for the busy beaver, the hardest-working Turing machine. *Scientific American*, 251(2):19–23, August 1984.
- [6] A. K. Dewdney. Computer recreations. *Scientific American*, 252(4):12–16, April 1985.
- [7] Solomon Feferman. Transfinite recursive progressions of axiomatic theories. *Journal of Symbolic Logic*, 27(3):259–316, 1962.
- [8] Solomon Feferman. *The Universal Turing Machine: A Half-Century Survey*, chapter Turing in the land of $\mathbf{O}(z)$, pages 113–147. Oxford University Press, Oxford, England, 1988.
- [9] Rona J. Kopp. The busy beaver problem. Mathematical sciences, State University of New York at Binghamton, 1981.
- [10] Grégory Lafitte. Gödel incompleteness revisited. In Bruno Durand, editor, *Proceedings of the First Symposium on Cellular Automata Journées Automates Cellulaires (JAC'2008)*, pages 74–89, Moscow, April 2008. MCCME Publishing House.
- [11] Shen Lin. *Computer Studies of Turing Machine Problems*. PhD thesis, The Ohio State University, Columbus (Ohio), 1963.
- [12] Grégory Lafitte and Christophe Papazian. The fabric of small Turing machines. In S. Barry Cooper, Thomas F. Kent, Benedikt Löwe, and Andrea Sorbi, editors, *Computation and Logic in the Real World, Third*

- Conference of Computability in Europe, CiE 2007*, Quaderni del Dipartimento di Scienze Matematiche e Informatiche "Roberto Magari". University of Siena, June 2007.
- [13] Shen Lin and Tibor Radó. Computer studies of Turing machine problems. *Journal of the Association for Computing Machinery*, 12(2):196–212, April 1965.
 - [14] Ming Li and Paul Vitányi. *An introduction to Kolmogorov complexity and its applications*. Springer Verlag, New York, 1997.
 - [15] H. Marxen and J. Buntrock. Attacking the busy beaver 5. *Bulletin of the EATCS*, 40:247–251, 1990.
 - [16] Rona Machlin and Quentin F. Stout. The complex behavior of simple machines. *Physica D. Nonlinear Phenomena*, 42:85–98, 1990.
 - [17] Piergiorgio Odifreddi. *Classical Recursion Theory*. North Holland Publishing, 1989.
 - [18] Piergiorgio Odifreddi. *Classical Recursion Theory*, volume Vol. II. North Holland Publishing, 1999.
 - [19] Tibor Radó. On non-computable functions. *Bell System Technical Journal*, 41:877–884, May 1962.
 - [20] Hartley Rogers. Gödel numberings of partial recursive functions. *Journal of Symbolic Logic*, 23:331–341, 1958.
 - [21] Hartley Rogers. *The Theory of Recursive Functions and Effective Computability*. MIT Press, 1967.
 - [22] Alexander Shen. Algorithmic information theory and Kolmogorov complexity. Unpublished lecture notes.
 - [23] R. M. Smullyan. *Recursion Theory for Metamathematics*. Oxford University Press, New York, 1993.
 - [24] A. M. Turing. Systems of logic based on ordinals. *Proceedings of the London Mathematical Society*, 45:161–228, 1939.
 - [25] Nikolai Vereshchagin and Alexander Shen. *Computable Functions*. American Mathematical Society, 2003.